Assimilation of SMOS Retrievals in the Land Information System

Clay B. Blankenship
Jonathan L. Case
Bradley T. Zavodsky
William L. Crosson

ABSTRACT

The Soil Moisture and Ocean Salinity (SMOS) satellite provides retrievals of soil moisture in the upper 5 cm with a 30-50 km resolution and a mission accuracy requirement of 0.04 cm^3 cm^{-3}. These observations can be used to improve land surface model soil moisture states through data assimilation. In this paper, SMOS soil moisture retrievals are assimilated into the Noah land surface model via an Ensemble Kalman Filter within the NASA Land Information System. Bias correction is implemented using Cumulative Distribution Function (CDF) matching, with points aggregated by either land cover or soil type to reduce sampling error in generating the CDFs. An experiment was run for the warm season of 2011 to test SMOS data assimilation and to compare assimilation methods. Verification of soil moisture analyses in the 0-10 cm upper layer and root zone (0-1 m) was conducted using *in situ* measurements from several observing networks in the central and southeastern United States. This experiment showed that SMOS data assimilation significantly increased the anomaly correlation of Noah soil moisture with station measurements from 0.45 to 0.57 in the 0-10 cm layer. Time series at specific stations demonstrate the ability of SMOS DA to increase the dynamic range of soil moisture in a manner consistent with station measurements. Among the bias correction methods, the correction based on soil type performed best at bias reduction but also reduced correlations. The vegetation-based correction did not produce any significant differences compared to using a simple uniform correction curve.

1 INTRODUCTION

Soil moisture is a quantity of direct interest in such diverse fields as agriculture, public health, water management, ecology, drought/flood monitoring and forecasting, and military operations. It also impacts weather processes over land by governing the available moisture for evapotranspiration and surface sensible and latent heat fluxes. These quantities in turn affect boundary layer growth and destabilization. During warm-season months, diurnal heating and convective initiation depend strongly on evapotranspiration and available boundary layer moisture, which are substantially affected by soil moisture content. Therefore, accurate initialization of the land surface state in numerical weather prediction (NWP) models is important for depicting the exchange of heat and moisture at the land surface. Previous studies have demonstrated the impact of land surface initialization on NWP forecasts and its influence on such variables as planetary boundary layer height and cloud cover [1]–[12].

Land surface models (LSMs) provide a valuable tool for all of these applications, enabling the simulation of soil moisture and temperature (state variables) based on precipitation, surface temperature and humidity, and radiation inputs (forcing variables) as well as properties of the soil and land cover. Data assimilation (DA) can be used to update LSMs by combining model states (background) with observations of state variables or related quantities to provide an updated model analysis superior to either data source used alone [13]. This allows LSMs to benefit from a combination of observations of both precipitation (a forcing variable) and soil moisture (a state variable) [14].

Satellite observations are particularly well suited for updating models, since they provide global coverage with frequent revisit times (usually once or twice daily for a polar orbiting satellite). Due to the influence of dielectric constant on emissivity, low-frequency (window

channel) microwave brightness temperatures are sensitive to the liquid water content near the soil surface (and in the vegetation canopy) [15], [16]. Microwave remote sensing data, either in the form of directly observed radiances (or equivalently, brightness temperatures), or retrieved products, have successfully been used to update the soil moisture state in LSMs. The first instruments used extensively for this purpose used channels in the 6-10 GHz range, beginning with the Scanning Multichannel Microwave Radiometer (SMMR) [17]–[19], followed by the Advanced Microwave Scanning Radiometer (AMSR) [20], [21], Tropical Rainfall Measurement Mission (TRMM) Microwave Imager (TMI) [12], and Microwave Radiation Imager (MWRI) [22]. Retrievals from an active sensor, the Advanced Scatterometer (ASCAT), have also been assimilated [20], [21], [23], and are being used operationally at the United Kingdom Met Office [24] and the European Centre for Medium-Range Weather Forecasts (ECMWF).

The era of operational L-band instruments began with the 2009 launch of the Soil Moisture and Ocean Salinity (SMOS) mission [25], [26] by the European Space Agency (ESA). The SMOS radiometer, Microwave Imaging Radiometer with Aperture Synthesis (MIRAS), uses a dual polarized synthetic aperture and senses radiation at 1.4 GHz. This lower frequency gives it the advantages of greater accuracy (0.04 cm^3/cm^3), better penetration of vegetation canopies, and a greater soil penetration depth (up to 5 cm) [27], at the cost of a larger horizontal footprint (35-50 km) [28]. Assimilation of SMOS retrievals [29], [30] and brightness temperatures [23], [31] is the focus of several current efforts, including operational brightness temperature monitoring at ECMWF [32] and Environment Canada [33].

Another L-band mission, NASA's Soil Moisture Active Passive (SMAP) [27] was launched in January 2015. SMAP includes a 1.41 GHz radiometer and a 1.2 GHz synthetic aperture radar. Its combined (active/passive) soil moisture retrieval product was intended to

provide a similar accuracy to SMOS with a higher resolution (9 km) due to the addition of the radar [34]. Unfortunately, the SMAP radar failed a few months into the checkout period [35], but it is hoped that this design could be the basis of future observing systems.

The goal of this research is to improve the accuracy of soil moisture analyses in an LSM by assimilating SMOS retrievals. These soil moisture fields may be used directly in applications such as agriculture, drought monitoring, and flood prediction. In this paper, we evaluate the impact of assimilating SMOS soil moisture observations into the Noah LSM (Section 2.3). These investigations are conducted within the framework of the Land Information System (LIS; Section 2.2). This research specifically focuses on the use of SMOS data in a widely used LSM based on a configuration (Section 4.1) used by forecasters for operational decision support.

Both direct radiance (or brightness temperature) assimilation [36]–[38] and retrieval assimilation [18], [19], [21] are commonly used for land surface assimilation in an effort to improve soil moisture analyses. Direct radiance assimilation offers some theoretical advantages such as a more accurate specification of error and the elimination of inconsistencies and/or correlated errors between retrievals and background [39]–[41]. However, there are some advantages to retrieval assimilation for land surface modeling and for SMOS data in particular: 1) the assimilated product may be monitored in model space in readily understandable soil moisture units; 2) the SMOS retrieval is a complex product utilizing overlapping observations at different incidence angles [42], which is not easily implemented into an assimilation system; 3) the SMOS sensor has shown substantial biases that are not well understood [42], [43], making bias correction in radiance space more problematic; and 4) the assimilation of bias-corrected retrievals is easily transferred to new sensors because the rescaled retrievals will have similar distributions even though the observing systems are quite different. The assimilation of SMOS

retrievals will therefore expedite the operational utilization of data from the recently launched SMAP mission.

2 MODELS AND OBSERVATIONS

2.1 SMOS

We assimilate soil moisture retrieved from the SMOS User Data Product Level 2 (SMUDP2) data [44], obtained from the ESA Centre Aval de Traitement des Donées (CATDS). This product has a resolution of 30-50 km with a mission accuracy requirement of 0.04 cm^3/cm^3 [25]. Validation experiments have shown results close to the target accuracy [45], with RMSE between 0.043 (morning) and 0.047 cm^3/cm^3 (evening) reported over sites in the U.S. [46]. The retrieval algorithm [45] uses multiple views of the same scene at different incidence angles and two polarizations. A Bayesian cost function is applied to find the soil moisture content and vegetation opacity that best fits the observed brightness temperatures and *a priori* parameter values. The algorithm uses static databases for soil texture, topography, and other properties, including the ECOCLIMAP land use database. Auxiliary dynamic data includes weather parameters such as snow cover, temperature, and rain from ECMWF as well as vegetation optical thickness based on Leaf Area Index (LAI). The algorithm accounts for the presence of subpixel inhomogeneity based on ECOCLIMAP, including barren (rocky) and urban areas, forests, and bodies of water.

The performance of the algorithm varies depending on the presence of vegetation and other factors. Due to canopy effects, performance is worse in dense vegetation. In extreme cases such as Amazonia, soil properties cannot be detected; over moderately dense forests such as boreal forest in Canada, RMSE ranges from 0.15 to 0.18 cm^3/cm^3 [47]. Significant interference from radio frequency interference (RFI) is most prevalent in parts of Europe and

Asia [48]. Scenes with high topographic index are also excluded due to the difficulty of simulating the surface topography.

2.2 NASA Land Information System

The NASA LIS is a land surface modeling framework developed at NASA-Goddard Space Flight Center that integrates community LSMs with forcing, parameter, and observational datasets within a high-performance computing environment [49], [50]. It includes a DA capability through an Ensemble Kalman Filter (EnKF). To facilitate intercomparisons, users may select LSMs, forcing data sources, land cover and soil type data sources, and other parameters from a variety of datasets. The architecture is modular to enable users to add new models and datasets. We have adapted LIS to assimilate the SMOS retrieved soil moisture product into the Noah LSM, as described in Section 3.

2.3 Noah Land Surface Model

We employ version 3.2 of the Noah LSM. Developed jointly by the National Center for Atmospheric Research (NCAR) and the National Centers for Environmental Prediction (NCEP), the Noah LSM is a unified code for research and operations. Temperature and water content are modeled in four soil layers with thicknesses of 0.1, 0.3, 0.6, and 1 m, as well as in the vegetation canopy and a possible snow layer [4], [51], [52]. The model includes formulations for evapotranspiration, soil drainage and diffusion, and runoff. The Noah LSM also predicts soil ice, fractional snow cover effects, has specialized urban treatment, and considers surface emissivity properties. When coupled to an NWP model, the scheme provides sensible and latent heat fluxes to the atmospheric boundary layer. [53] provides a comprehensive summary of the formulation and physical components comprising the Noah LSM as run at the NCEP Environmental Modeling Center.

2.4 Ensemble Kalman Filter

The assimilation of SMOS retrievals is implemented using the EnKF [54][55] assimilation system included in LIS [56]. A Kalman Filter combines a forecast (background) with observations (soil moisture retrievals, in this case) to generate an improved estimate of a model variable. In an EnKF, an ensemble of model runs is used to represent the model state and its associated uncertainty. Here, the ensemble is generated with perturbations to represent errors in the state, forcing, and observations. At each assimilation step, the new state of each ensemble member is computed as a weighted average of the observations (retrieved soil moistures) and the model background. The weighting is determined by the relative size of the background and observation errors, with the background error being determined by the ensemble covariance. The observation operator is simply a rescaling (described in Section 3.1) of the top layer of the soil moisture state (0-10 cm in Noah). Because of this, only the top layer is adjusted directly, but deeper layers of soil moisture can change slightly in the DA step due to background covariances. A more significant impact (over time) on the deeper layers can occur resulting from the modeled physical processes of drainage and diffusion [57]. In the present LIS architecture, observations are assimilated separately at each cell, so there are no horizontal background correlations.

2.5 Demonstration of Assimilation

An illustration of the assimilation process is shown in Figure 1 for a scenario with an unusually strong signal observed at 0000 UTC 1 April 2013. Prior to any assimilation (Figure 1a), the model had moderate-to-high soil moisture over most of the domain. The SMOS satellite, whose swath traverses the western half of the domain, indicates extremely wet conditions over much of the lower Mississippi Valley (Figure 1b), whereas the model background underestimated the soil moisture where irrigation had occurred in this region. The analysis

(Figure 1c) combines features of both fields but clearly depicts the area of enhanced soil moisture from the SMOS retrievals. Figure 1d shows irrigated areas from the Food and Agriculture Organization (FAO) of the United Nations [58], [59], largely coinciding with the area of elevated soil moisture shown in the analysis. Since no irrigation data source was used in the LIS run, the irrigation signal is not present in the model background, but the wet area is incorporated into the model through the assimilation process. The large amount of available moisture and resulting land-atmosphere interaction in this part of the southeastern U.S. during the spring and summer may have significant impacts on the boundary layer and convective potential. Hence, better depiction of these processes in a numerical model could potentially improve weather forecasts in these situations.

3 IMPLEMENTATION

3.1 Bias Correction

3.1.1 Cumulative Distribution Function Matching

Assimilation systems generally assume that observations are unbiased relative to the model [60]. There are many sources of error in model and observations (whether using radiance or retrieval assimilation), including measurement (and retrieval) errors, forcing errors, and uncertainties in soil and vegetation parameters. When assimilating data, there may be systematic errors due to the discrepancy between the sensor's detection depth and the modeled layer depth, due to sharp vertical gradients near the surface [61]. Moreover, there are representativeness errors due to the vertical resolution of the model and the variability in the penetration depth of the sensor. Bias correction techniques are employed in most if not all satellite DA efforts (but rarely for *in situ* observations) in order to adjust the observations to fit the model climatology [12], [62].

One of the most common bias correction techniques in land DA is Cumulative Distribution Function (CDF) matching [17], in which an observation is corrected to a model-equivalent value. CDF matching is often applied by building up model and observation CDFs at each grid cell independently [17], but a weakness of this approach is that it needs a long training period to obtain a reasonable sample size to establish the correction curves. This long training can be reduced by using a larger search radius around each point to build up the *a priori* distributions [17], [21]. This method assumes ergodicity and substitutes spatial variability for temporal variability. [17] found a period of one year was sufficient to establish CDF curves for a 2-degree search radius. However, with soil moisture satellite missions, such as SMOS and SMAP only scheduled for 3-5 years of operations, even a one-year bias-correction period does not allow for use of the data early in the mission lifetime. The use of multi-year training periods also precludes the use of corrections to account for seasonal biases or instrument drift [63]. While most existing soil moisture bias correction efforts correct the observations to match the model distribution at each point, other approaches are possible. For example, [64] suggested using a single first-order correction in global assimilation efforts.

A strength of satellite observations is that they are physically based, globally consistent measurements. In contrast to pointwise methods, a looser constraint (e.g. matching the CDF in the aggregate rather than at each grid cell) can use the physically consistent satellite observations to correct for spatially varying biases in the model, perhaps due to heterogeneous errors in the forcing data, or to unmodeled processes such as irrigation. This approach allows observational datasets of both precipitation and soil moisture to influence the model climatology.

3.1.2 Variations Tested

We compare three bias correction methods based on spatial aggregation of a large number of points. The simplest method (BC1), a uniform CDF-matching correction applied to all points, serves as a baseline for testing the more complex methods. Figure a shows CDFs for observations (SMOS retrievals) and coincident model values taken from March to September in 2013. When the correction is applied, each retrieved soil moisture values is converted to an equivalent value in the model climatology based on the two CDFs. The remaining two methods are based on shared physical properties. Results from all three bias correction methods are compared to results obtained from DA with no bias correction, and to an open loop (OPL) control run.

From Figure 2a, we observe that the observations are drier in general, and also have a larger dynamic range. The Level 2 soil moisture product, while normally interpreted as volumetric soil moisture, has a range that varies up to 1 and is not limited by the soil porosity value. This was an intentional choice by ESA to preserve the signal in flooded areas [44]. To correct SMOS retrievals into a model-appropriate value, the CDF correction converts the Level 2 data value (with a maximum value of 1) to a model-equivalent value of volumetric soil moisture (with a maximum value equal to the soil porosity).

Several sources of bias vary systematically with physical properties of the surface and vegetation. Some of these bias sources are related to the radiative signal, such as the thickness of vegetation, affecting the microwave sensitivity to the surface layer [12]. Others, such as wilting point and field capacity [12, 50], are hydrological. We have hypothesized that the bias characteristics are strongly dependent on either land use type or soil type, and developed two variants of the CDF-matching correction [17] aggregating points with similar soil and land use

classifications. We previously applied a similar approach to the assimilation of AMSR-E soil moisture retrievals based on land use only [66].

The soil classes used were initially assigned according to the State Soil Geographic (STATSGO) [67] database. To reduce sampling error for classes that had relatively few points, these classifications were aggregated into groups with physically similar types as well as similar CDFs (Table 1), resulting in seven groups. For example, STATSGO classes 2 and 3 (Loamy Sand and Sandy Loam) were combined into the "Sandy Loam Group". Each of the seven groups has a correction curve derived based on the CDF of coincident observations (SMOS retrievals) and model values (Figure 2b). Following a similar procedure, a separate correction based on land cover was derived using the International Geosphere-Biosphere Programme (IGBP) Moderate Resolution Imaging Spectroradiometer (MODIS) land use classification [68]. Classes were again aggregated based on geophysical similarities as well as similarity in CDFs (Table 2), resulting in six categories with separate correction curves (Figure 2c).

Grouping points by land cover or soil classification serves the same purpose as the search radius in [17], reducing sampling error for building the CDFs by combining many spatial points. This looser constraint enables the observations to adjust the climatological distributions geographically. However, it might give poorer results if biases depend strongly on other factors. In light of this, bias correction maps should be monitored to understand shifts in spatial soil moisture patterns and determine whether the changes are correcting spatial biases in the model.

Other variations on this method could be applied to limit the aggregated points to a certain radius, stratify by soil type rather than land cover class, and/or allow the correction to vary seasonally as in [21]; however the objective of this paper is to demonstrate an initial methodology for assimilation of SMOS data using a basic technique. By implementing a method

that only needs a few months of satellite records, we plan to establish a technique that can be quickly applied for SMAP DA.

The correction curves used in this study were derived from coincident observation and modeled data from an independent monitoring (non-assimilating) run conducted for March to September 2013. One particular challenge for generating a CDF over the southeast U.S. is that the domain includes regions around the lower Mississippi Valley with extensive irrigation (primarily for rice) in the spring and early summer months that is not captured in the forcing data, but is observed by the SMOS sensor. As this methodology allows model distributions to adjust in order to fit the observations, it was important to eliminate points with a known significant model bias from the training set. Therefore, an area around the irrigated region, shown as the box in Figure 3, was excluded from the CDF database prior to June 1.

3.2 Quality Control

Quality control is applied after eliminating overlapping observations and incorporates both model fields and SMOS data flags. The SMOS data flags [69] used to reject retrievals include Data Quality Index (DQX>0.06, [70]), Global Quality Index (GQX>10), RFI Probability (RFI_Prob>0.25), and Soil Moisture (negative values indicating). Where the precipitation (from the model forcing data) is above 1 mm/hr, observations are rejected, since emission from liquid water in the atmosphere obscures the surface signal. Observations are also rejected where the surface is frozen and/or snow covered since the algorithm is designed to sense liquid water.

4 DATA ASSIMILATION EXPERIMENT

4.1 Study Area

The model domain (Figure 1) includes the central and southeastern United States with a ~3-km grid spacing. It includes a variety of temperate land cover types listed in Table 2. This

domain is used by the NASA Short-term Prediction Research and Transition (SPoRT) Center to provide near-realtime LSM output to National Weather Service forecasters. This product has been used by operational forecasters as input for drought monitoring [71] and for situational awareness for flood potential [72]. Performing the experiments in this domain using a similar LIS configuration allows us to test the impact of SMOS data in an environment that could be used for decision support. This scale is consistent with the 1-3 km resolutions commonly used regional and national (CONUS) LSM and NWP configurations, and enables the depiction of gradients in soil moisture on sub-county scales that can be used to better depict hydrological and meteorological features associated with localized flooding and drought. This grid spacing is also fine enough to allow convection in NWP models [73], which means that the current domain can be used for future coupled land/atmosphere examining the impact of initial soil moisture conditions on convective initiation.

With a grid spacing of ~3 km, the model resolution is much finer than the retrieval resolution. Consequently, LIS has been configured to apply each observation to multiple points within the observation field of view (FOV). In the case of overlapping FOVs, only the nearest observation is retained for each grid cell. The maximum radius was set at 18 km, considering the range of FOV sizes (30-50 km diameter) and the observation spacing. Some disaggregation is achieved through the land-cover-based bias correction. A more idealized approach might include a forward operator that averages model values over all grid points within the FOV. Future data products may be available on a resolution similar to this grid spacing, as demonstrated by the 3-km resolution of the short-lived SMAP radar.

4.2 Experiment Design and Model Configuration

We tested the impact of assimilating SMOS retrievals using ensemble runs of the Noah LSM within LIS for the period from 1 Feb. to 1 Oct. 2011. Additionally, the bias correction methodology was evaluated by separately testing assimilation with three versions of the bias correction as well as one simulation with assimilation but no bias correction. The validation period was begun on 1 March to allow the results to reflect the impact of repeated DA in a cycling framework. The model runs (Table 3) were: 1) a control run (open loop, OPL) with perturbations but no assimilation, 2) a SMOS DA run with no bias correction (NoBC), 3) SMOS DA with a uniform (spatially invariant) bias correction scheme (BC1), 4) SMOS DA with a soil-type-dependent bias correction (BCS), and 5) SMOS DA with a bias correction dependent on vegetation/land cover (BCV).

All simulations consisted of 32-member ensemble runs. These were initialized with a multi-year single member spin-up ending on 1 Feb. 2011, followed by a one-month open loop ensemble run to spin up the state and forcing perturbations. Perturbations were generated using the methodology and values of [74]. The forcing perturbations consisted of additive perturbations with σ=50 W m^{-2} to longwave radiation, and multiplicative perturbations for shortwave radiation (σ=0.3) and precipitation (σ=0.5), where the shortwave perturbations were negatively correlated with the other two. For the DA simulations, the observation perturbation was a random normal variable with a standard deviation of 0.04 cm^3 cm^{-3}, based on the stated SMOS design accuracy [25]. It is expected that this value could be optimized by continued experimentation. Further refinement may be possible by allowing the SMOS uncertainty to vary in space or time, since the data do include uncertainty information and previous studies have shown that the retrieval accuracy is variable [75], [76].

Model runs were forced with temperature, humidity, winds, and incident radiation at the surface from the North American Land Data Assimilation System-2 (NLDAS-2) at 1/8-degree resolution [77] and 4-km gridded precipitation from Stage IV combined radar/rain gauge analyses [78]. Land use categories were set according to the IGBP land-use classification [68] as applied to the MODIS instrument [79]. Soil types were classified according to the State Soil Geographic (STATSGO) [67] database. Static and dynamic land surface fields were masked based on the IGBP/MODIS land-use classes. Static properties including porosity, saturated suction, saturated hydraulic conductivity, field capacity (with minor changes), and wilting point are soil type-dependent and were set from lookup table values published by [4] based on measurements by [80]. Other static fields such as minimum canopy resistance are assigned according to the land use category. Green vegetation fraction (GVF) was derived from daily real-time MODIS data [81]. The Noah LSM determines albedo, roughness length, and LAI from GVF based on a linear scaling and ranges defined in the land use lookup table.

The Noah LSM was run with a model timestep of 30 minutes. Retrievals are available up to twice daily at each location, with some missing data due to quality control and gaps between swaths. LIS was configured to assimilate data on a 6-hour time step. For this domain, all the observations were assimilated during the 0000 and 1200 UTC cycles (roughly 6 a.m. and 6 p.m. local time) due to the sun-synchronous orbit of SMOS. There was no time interpolation or other temporal correction, rather each observation that occurred during the 6-hour window was applied simultaneously at either 0000 or 1200 UTC. This was not expected to be a significant source of error since the observations occur within about 2 hours of the assimilation time and soil moisture generally varies slowly apart from rain events.

4.3 Validation Methodology

In situ measurements of soil moisture are a valuable independent and well-calibrated dataset for LSM validation. Validation of the model runs was conducted by comparing model soil moisture values for the 0-10 cm (surface layer) and 10-100 cm (root zone) layers from each simulation against *in situ* observations from the North American Soil Moisture Database (NASMD) [82] obtained from Texas A&M University. These station observations were used for validation only and were not used as inputs to the any of the model simulations. The model's upper layer (0-10 cm) soil moisture was validated against measurements from the corresponding depths (typically 5 cm). Additionally, a weighted average of available *in* situ measurements from 10-100 cm was used to validate the model's root zone soil moisture. To eliminate unreliable stations, stations with fewer than 150 days of observations and those that were negatively correlated with any model run (applying a 15-day moving average) were removed. After quality control, there remain 194 stations for validation of the surface layer, and 137 stations for the root zone.

Values for bias, root-mean square error (RMSE), unbiased RMSE (ubRMSE), and Pearson correlation were calculated at the chosen stations for the period 1 Apr. 2011 to 1 Oct. 2011 for each model simulation. Because comparisons of model grid-scale values to point-based field measurements is limited by representativeness errors, and due to possible mismatches between model and actual properties such as soil type and vegetation class, we expect the latter two metrics (which are insensitive to bias) to be the best indicators of DA performance. We hypothesize that SMOS assimilation will significantly improve correlation and ubRMSE. Due to the intrinsic biases between satellite and station data, we expect the magnitude of bias to increase for all DA runs compared to the OPL run. All three metrics can be used to judge the quality of

the bias correction methods, although we do not expect the differences to be large.

5 VALIDATION RESULTS

5.1 SMOS retrieval validation

As a preliminary step to model verification, we compared SMOS retrieved soil moistures to the field measurements from the NASMD. This was not intended to be a rigorous evaluation of SMOS accuracy but rather to give context to the model validation results. For a more complete validation of SMOS retrievals, the reader is referred to [46, 47, 83-85]. Bias, RMSE, ubRMSE, correlation, and anomaly correlation were calculated for collocated, quality-controlled SMOS and ground station observations in the domain within the experiment time period (1 Apr. 2011 to 1 Oct. 2011). From the stations passing quality control in Section 4.3, we chose those that had at least 20 satellite matches during this period, yielding a validation dataset of 182 stations.

Results from comparing SMOS retrieved soil moisture and ground station measurements are given in Table 4 and Figure 4. Table 4 summarizes the statistical results (bias, RMSE, ubRMSE, correlation, and anomaly correlation with their 95% confidence intervals) from the 182 stations with a total of 15,017 collocated observations. Statistical metrics were calculated using equal station weighting and the ubRMSE was calculated by subtracting the mean bias (constant term) over the whole dataset. The RMSE of 0.121 is three times as large as the accuracy goal of 0.04. However the ubRMSE, calculated by subtracting out the constant mean bias term, is 0.071, which is within a factor of two.

Histograms for the five evaluation metrics are presented in Figure 4, exhibiting a wide range of performance across stations. Note that discrepancies between these two soil moisture estimates can be attributed to a) measurement error of the ground station, b) errors in the

retrieval product including radiometric measurement error, and c) representativeness error. Of these error sources, the first should be low for well-calibrated stations whereas the other sources may vary significantly between stations (e.g., retrieval error is larger in dense vegetation, and representativeness error may be worse when the satellite FOV is heterogeneous). If we attribute the spread in the histograms chiefly to the variation in representativeness error, we can estimate the RMSE of the SMOS retrieval to be between 0.04 (best case) and 0.08 (mode of all values). Using similar reasoning, we can estimate the ubRMSE to range from 0.04 to 0.06. These estimates apply to a "best case" scenario where retrieval errors are minimized (e. g., for areas with low to moderate vegetation).

5.2 Experiment results

Figure 5 shows modeled and observed time series of upper layer (0-10 cm) and root zone (0-1 m) soil moisture from three selected stations (among those exhibiting the largest impact) in Missouri, Oklahoma, and Illinois. Vertical bars indicate rainfall (from the model forcing data), which causes a sharp increase in upper layer soil moisture, typically followed by an asymptotic drying behavior. Significant rain events generally coincide with increases in the station data (black curves). However, less significant rain events do not always appear in the station time series, indicating small errors in the forcing. For simplicity, values from only the OPL, NoBC, and BC1 model runs are plotted. The control run (OPL, in red) provides a baseline for the impact of uncorrected DA (NoBC, blue). The magnitude of bias corrections is generally small, seen by the separation of the green (BC1) and blue curves.

Both DA runs exhibit some added noise, particularly in the upper zone, which is most likely due to random error in the SMOS observations. Each of these stations shows an improvement in correlation due to assimilating SMOS data. DA increased the model dynamic

range for each case, better matching the station measurements. The DA runs are better able to capture the seasonal drying trend seen in both layers, with closer values to the station measurements both in the early spring when the soil is relatively wet and in the fall when the soil is drier. The DA runs also show an increased (improved) dynamic range on shorter scales in the upper zone such as the drying trend during July and after the rain in early August in panels a and c, and in the March and April drying trend in panel c. The limited dynamic range of the OPL runs for these cases may be an indication that some model parameters are not properly optimized, and parameter estimation methods (e.g. [86]–[88]) may improve the model behavior here. Nevertheless, these examples illustrate how DA of SMOS soil moisture retrievals can improve model estimates in the presence of model weaknesses.

Tables 5 and 6 summarize the results over all validation sites for the upper zone (0-10 cm) and root zone (0-1 m). All values were calculated using the Land surface Verification Toolkit (LVT) [89] after applying QC to select stations. In the upper zone, the effect of DA is seen most clearly in the impact on correlation scores. All DA runs show a significant increase in correlation compared to the control run (OPL). The overall impact on RMSE and ubRMSE is not significant, which is not surprising given that the model control run matches the stations better than the SMOS retrievals do. The NoBC run exhibits the highest correlation of all runs, 0.573, compared to 0.451 for OPL. Because the OPL bias (relative to the stations) is coincidentally very close to zero, it is not surprising that NoBC increased the magnitude of the bias, to a value of -0.026 cm^3 cm^{-3}. Since they are designed to correct toward the OPL model, all bias correction methods were expected to reduce the bias compared to NoBC, but only the BCS (soil-based correction) significantly reduced it, indicating that the soil type-dependent method is most effective at reducing bias based on previously calculated correction curves. However, the

BCS has a lower correlation than the other bias correction methods (BC1 and BCV). The unbiased RMSE had little variation, showing a small, statistically insignificant improvement due to DA. The two category-based corrections did not significantly reduce the unbiased RMSE as expected. A follow-up investigation should be conducted to determine if the correction curves shown in Figure 2 are consistent interannually and interseasonally. In the root zone, the bias exhibited some statistically significant changes, but due to intrinsic biases between the station measurements and model values, it is difficult to draw conclusions from these changes. The other metrics did not show statistical significance.

Maps indicating changes in correlation and unbiased RMSE at each station location are shown in Figure 6. In each map, blue colors indicate improvement and red indicates degradation. The maps indicate that the anomaly correlation improvements are widespread (particularly in the 0-10 cm layer) with a high degree of spatial coherence. Performance was best in the northern/western portions of the domain, and poorer in the southeastern quadrant. The reduction in unbiased RMSE, although statistically insignificant, is evident in the majority of locations for the upper layer. For the root zone, the changes are not consistently in one direction, thus showing no particular improvement.

6 SUMMARY

The NASA SPoRT Center has implemented the assimilation of soil moisture retrievals from the SMOS satellite within the NASA LIS. A bias correction methodology using a CDF matching technique, stratified by land cover class or soil type, has also been developed. Validation against the North American Soil Moisture Database shows that upper layer (0-10 cm) correlation scores are clearly improved for all DA experiments (three bias correction methods plus no bias correction) relative to the OPL baseline. A map of correlation improvements shows

that the effect is widespread. Time series at selected stations show that DA improves both short-term and seasonal trends, with larger dynamic ranges from spring to fall being evident at the selected stations in both upper and root zones. Unbiased RMSE reduction due to DA was not statistically significant although the map of changes showed the improvements were generally consistent. No significant changes were detected in the root zone, although correlations had a slight improvement due to DA as confirmed by a map.

Because the retrieval dynamic ranges and/or distributions can differ markedly from the model, bias correction is desirable. The soil-type-based correction was best at reducing bias compared to stations, but had slightly worse correlation scores than the other methods. The land-cover-based correction showed no significant benefits over the simple uniform correction when validating against stations. As an attempt to improve on these results, bias correction curves should be examined for different years and seasons to see how much variation they exhibit. Hybrid bias correction methods could also be explored, using a radius of influence as in [17] but limited to similar soil or land use type.

The study period was limited to one warm season over the southeastern U.S. Further investigation could reveal if the SMOS data has similar impact during other years and in other geographic regions. The experiment was conducted using forcing data from the robust Stage IV and NLDAS-2 datasets available over North America. Tests of the DA system in regions with poorer forcing data quality would be expected to show even stronger impact. Other improvements to the DA procedure could be implemented, such as 1) specification of the retrieved soil moisture uncertainty using satellite data quality indicators; 2) development of a forward operator that properly aggregates all pixels within the field of view; 3) ensuring

consistency between the ECOCLIMAP subpixel vegetation database used in the SMOS retrievals and the LSM land cover classification.

The SMOS assimilation will be implemented in SPoRT's near-real-time LIS modeling efforts and made available on the SPoRT web site (http://weather.msfc.nasa.gov/sport). SPoRT is also working to assimilate SMAP Level 2 soil moisture retrievals, based on experience gained from assimilating SMOS data. Because previous studies have demonstrated the impact of more accurate surface boundary and initial conditions on numerical weather forecasts, future efforts will investigate the impact of assimilating SMOS and SMAP data on regional NWP model runs initialized from LIS output, assessing the sensitivity of weather forecasts to SMOS and SMAP assimilation and also validating the forecasts against observations from ground stations, soundings and precipitation analyses.

ACKNOWLEDGEMENTS

This work is supported by the NASA Science Mission Directorate. SMOS data were supplied by the European Space Agency. We would like to thank Susan Moran, Vanessa Escobar, and Molly Brown for including us as SMAP Early Adopters. We are grateful to William Campbell of the Naval Research Laboratory for feedback about bias correction and to Yann Kerr and Arnaud Mialon of Centre d'Études Spatiales de la BIOsphère (CESBIO) for helpful discussions about the SMOS Level 2 product. Finally, we thank the reviewers for their helpful comments during the revision process.

REFERENCES

[1] R. D. Koster and M. J. Suarez, "Impact of Land Surface Initialization on Seasonal Precipitation and Temperature Prediction," *Journal of Hydrometeorology*, vol. 4, no. 2. pp. 408–423, 2003.

[2] R. D. Koster, M. J. Suarez, P. Liu, U. Jambor, A. Berg, M. Kistler, R. Reichle, M. Rodell, and J. Famiglietti, "Realistic Initialization of Land Surface States: Impacts on Subseasonal Forecast Skill," *J. Hydrometeorol.*, vol. 5, no. 6, pp. 1049–1063, 2004.

[3] F. Chen and R. Avissar, "Impact of Land-Surface Moisture Variability on Local Shallow Convective Cumulus and Precipitation in Large-Scale Models," *J. Appl. Meteorol.*, vol. 33, no. 12, pp. 1382–1401, Dec. 1994.

[4] F. Chen and J. Dudhia, "Coupling an Advanced Land Surface–Hydrology Model with the Penn State–NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity," *Mon. Weather Rev.*, vol. 129, no. 4, pp. 569–585, Apr. 2001.

[5] E. A. B. Eltahir, "A Soil Moisture-Rainfall Feedback Mechanism: 1. Theory and observations," *Water Resour. Res.*, vol. 34, no. 4, pp. 765–776, Apr. 1998.

[6] K. L. Findell and E. A. B. Eltahir, "Atmospheric Controls on Soil Moisture–Boundary Layer Interactions. Part II: Feedbacks within the Continental United States," *J. Hydrometeorol.*, vol. 4, no. 3, pp. 570–583, Jun. 2003.

[7] S. B. Trier, F. Chen, and K. W. Manning, "A Study of Convection Initiation in a Mesoscale Model Using High-Resolution Land Surface Initial Conditions," *Mon. Weather Rev.*, vol. 132, no. 12, pp. 2954–2976, Dec. 2004.

[8] S. B. Trier, F. Chen, K. W. Manning, M. A. LeMone, and C. A. Davis, "Sensitivity of the PBL and Precipitation in 12-Day Simulations of Warm-Season Convection Using Different Land Surface Models and Soil Wetness Conditions," *Mon. Weather Rev.*, vol. 136, no. 7, pp. 2321–2343, Jul. 2008.

[9] M. A. LeMone, F. Chen, J. G. Alfieri, R. H. Cuenca, Y. Hagimoto, P. Blanken, D. Niyogi, S. Kang, K. Davis, and R. L. Grossman, "NCAR/CU Surface, Soil, and Vegetation Observations during the International H 2 O Project 2002 Field Campaign," *Bull. Am. Meteorol. Soc.*, vol. 88, no. 1, pp. 65–81, Jan. 2007.

[10] J. L. Case, W. L. Crosson, S. V. Kumar, W. M. Lapenta, and C. D. Peters-Lidard, "Impacts of High-Resolution Land Surface Initialization on Regional Sensible Weather Forecasts from the WRF Model," *J. Hydrometeorol.*, vol. 9, no. 6, pp. 1249–1266, 2008.

[11] J. L. Case, S. V. Kumar, J. Srikishen, and G. J. Jedlovec, "Improving Numerical Weather Predictions of Summertime Precipitation over the Southeastern United States through a High-Resolution Initialization of the Surface State," *Weather Forecast.*, vol. 26, no. 6, pp. 785–807, Dec. 2011.

[12] M. Drusch, "Initializing numerical weather prediction models with satellite-derived surface soil moisture: Data assimilation experiments with ECMWF's integrated forecast system and the TMI soil moisture data set," *J. Geophys. Res. Atmos.*, vol. 112, no. 3, p. D03102, Feb. 2007.

[13] R. H. Reichle and R. D. Koster, "Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model," *Geophys. Res. Lett.*, vol. 32, no. 2, p. L02404, 2005.

[14] Q. Liu, R. H. Reichle, R. Bindlish, M. H. Cosh, W. T. Crow, R. de Jeu, G. J. M. De Lannoy, G. J. Huffman, and T. J. Jackson, "The Contributions of Precipitation and Soil Moisture Observations to the Skill of Soil Moisture Estimates in a Land Data Assimilation System," *J. Hydrometeorol.*, vol. 12, no. 5, pp. 750–765, Oct. 2011.

[15] T. J. Schmugge and T. J. Jackson, "Passive microwave remote sensing system for soil moisture: some supporting research," *IEEE Trans. Geosci. Remote Sens.*, vol. 27, no. 2, pp. 225–235, Mar. 1989.

[16] M. Owe and A. A. Van de Griend, "Comparison of soil moisture penetration depths for several bare soils at two microwave frequencies and implications for remote sensing," *Water Resour. Res.*, vol. 34, no. 9, pp. 2319–2327, Sep. 1998.

[17] R. H. Reichle and R. D. Koster, "Bias reduction in short records of satellite soil moisture," *Geophys. Res. Lett.*, vol. 31, no. 19, L19501, Oct. 2004.

[18] M. Drusch, E. F. Wood, and H. Gao, "Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture," *Geophys. Res. Lett.*, vol. 32, no. 15, L15403, Aug. 2005.

[19] W. Ni-Meister, P. R. Houser, and J. P. Walker, "Soil moisture initialization for climate prediction: Assimilation of scanning multifrequency microwave radiometer soil moisture data into a land surface model," *J. Geophys. Res. Atmos.*, vol. 111, no. 20, D20102, Oct. 2006.

[20] C. S. Draper, R. H. Reichle, G. J. M. De Lannoy, and Q. Liu, "Assimilation of passive and active microwave soil moisture retrievals," *Geophys. Res. Lett.*, vol. 39, no. 4, L04401, Feb. 2012.

[21] C. S. Draper, J. F. Mahfouf, and J. P. Walker, "An EKF assimilation of AMSR-E soil moisture into the ISBA land surface scheme," *J. Geophys. Res. Atmos.*, vol. 114, no. 20, D20104, Oct. 2009.

[22] Q. Lu, "Initial evaluation and assimilation of FY-3A atmospheric sounding data in the ECMWF System," *Sci. China Earth Sci.*, vol. 54, no. 10, pp. 1453–1457, Jul. 2011.

[23] P. de Rosnay, M. Drusch, G. Balsamo, L. Isaksen, and C. Albergel, "Extended Kalman Filter soil moisture analysis in the IFS," *ECMWF Spring Newsletter n127*, pp. 12–16, 2011.

[24] I. Dharssi, K. J. Bovis, B. Macpherson, and C. P. Jones, "Operational assimilation of ASCAT surface soil wetness at the Met Office," *Hydrol. Earth Syst. Sci.*, vol. 15, no. 8, pp. 2729–2746, Aug. 2011.

[25] Y. H. Kerr, P. Waldteufel, J. P. Wigneron, J. M. Martinuzzi, J. Font, and M. Berger, "Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 8, pp. 1729–1735, 2001.

[26] Y. H. Kerr, P. Waldteufel, J.-P. Wigneron, S. Delwart, F. Cabot, J. Boutin, M.-J. Escorihuela, J. Font, N. Reul, C. Gruhier, S. E. Juglea, M. R. Drinkwater, A. Hahne, M. Martín-Neira, and S. Mecklenburg, "The SMOS Mission: New Tool for Monitoring Key Elements ofthe Global Water Cycle," *Proc. IEEE*, vol. 98, no. 5, pp. 666–687, May 2010.

[27] D. Entekhabi, E. G. Njoku, P. E. O'Neill, K. H. Kellogg, W. T. Crow, W. N. Edelstein, J. K. Entin, S. D. Goodman, T. J. Jackson, J. Johnson, J. Kimball, J. R. Piepmeier, R. D. Koster, N. Martin, K. C. McDonald, M. Moghaddam, S. Moran, R. Reichle, J. C. Shi, M. W. Spencer, S. W. Thurman, L. Tsang, and J. Van Zyl, "The soil moisture active passive (SMAP) mission," *Proc. IEEE*, vol. 98, no. 5, pp. 704–716, May 2010.

[28] T. J. Jackson, J. Schmugge, and E. T. Engman, "Remote sensing applications to hydrology: soil moisture," *Hydrol. Sci. J.*, vol. 41, no. 4, pp. 517–530, 1996.

[29] M.-E. Ridler, H. Madsen, S. Stisen, S. Bircher, and R. Fensholt, "Assimilation of SMOS-derived soil moisture in a fully integrated hydrological and soil-vegetation-atmosphere transfer model in Western Denmark," *Water Resour. Res.*, vol. 50, no. 11, pp. 8962–8981, Nov. 2014.

[30] G. Dumedah and J. P. Walker, "Intercomparison of the JULES and CABLE land surface models through assimilation of remotely sensed soil moisture in southeast Australia," *Adv. Water Resour.*, vol. 74, pp. 231–244, Dec. 2014.

[31] R. H. Reichle, G. J. M. De Lannoy, W. T. Crow, R. D. Koster, and J. Kimball, "Assimilation of Smos Observations to Generate a Prototype SMAP Level 4 Surface and Root-Zone Soil Moisture Product," in *IEEE Geoscience and Remote Sensing Symposium*, 2012.

[32] J. Munoz Sabater, A. Fouilloux, and P. de Rosnay, "Technical Implementation of SMOS Data in the ECMWF Integrated Forecasting System," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 252–256, Mar. 2012.

[33] B. Bilodeau, M. L. Carrera, D. Charpentier, M. Lépine, B. Brasnett, and S. Bélair, "Towards the operational implementation of the Canadian Land Data Assimilation System," in *6th WMO Symposium on Data Assimilation*, 2013.

[34] N. N. Das, D. Entekhabi, and E. G. Njoku, "An algorithm for merging SMAP radiometer and radar data for high-resolution soil-moisture retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1504–1512, May 2011.

[35] S. P. Neeck, "The NASA Earth Science Flight Program: an update," in *SPIE Remote Sensing*, 2015, p. 963907.

[36] W. L. Crosson, C. A. Laymon, R. Inguva, and M. P. Schamschula, "Assimilating remote sensing data in a surface flux-soil moisture model," *Hydrol. Process.*, vol. 16, no. 8, pp. 1645–1662, 2002.

[37] A. S. Jones, T. Vukićević, and T. H. Vonder Haar, "A Microwave Satellite Observational Operator for Variational Data Assimilation of Soil Moisture," *J. Hydrometeorol.*, vol. 5, no. 1, pp. 213–229, Feb. 2004.

[38] G. J. M. De Lannoy, R. H. Reichle, and V. R. N. Pauwels, "Global Calibration of the GEOS-5 L-Band Microwave Radiative Transfer Model over Nonfrozen Land Using SMOS Observations," *J. Hydrometeorol.*, vol. 14, no. 3, pp. 765–785, 2013.

[39] J. C. Derber and W.-S. Wu, "The Use of TOVS Cloud-Cleared Radiances in the NCEP SSI Analysis System," *Mon. Weather Rev.*, vol. 126, no. 8, pp. 2287–2299, Aug. 1998.

[40] J. Joiner and D. P. Dee, "An error analysis of radiance and suboptimal retrieval assimilation," *Q. J. R. Meteorol. Soc.*, vol. 126, no. 565, pp. 1495–1514, Jul. 2007.

[41] R. H. Reichle, "Data assimilation methods in the Earth sciences," *Adv. Water Resour.*, vol. 31, no. 11, pp. 1411–1418, Nov. 2008.

[42] C. Montzka, H. R. Bogena, L. Weihermuller, F. Jonard, C. Bouzinac, J. Kainulainen, J. E. Balling, A. Loew, J. T. dall'Amico, E. Rouhe, J. Vanderborght, and H. Vereecken, "Brightness Temperature and Soil Moisture Validation at Different Scales During the SMOS Validation Campaign in the Rur and Erft Catchments, Germany," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1728–1743, Mar. 2013.

[43] G. J. M. De Lannoy, R. H. Reichle, and V. R. N. Pauwels, "Global Calibration of the GEOS-5 L-Band Microwave Radiative Transfer Model over Nonfrozen Land Using SMOS Observations," *J. Hydrometeorol.*, vol. 14, no. 3, pp. 765–785, Jun. 2013.

[44] European Space Agency, "Algorithm Theoretical Basis Document (ATBD) for the SMOS Level 2 Soil Moisture Processor Development Continuation Project", version 3.6, 2011.

[45] Y. H. Kerr, P. Waldteufel, P. Richaume, J. P. Wigneron, P. Ferrazzoli, A. Mahmoodi, A. Al Bitar, F. Cabot, C. Gruhier, S. E. Juglea, D. Leroux, A. Mialon, and S. Delwart, "The SMOS Soil Moisture Retrieval Algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1384–1403, May 2012.

[46] T. J. Jackson, R. Bindlish, M. H. Cosh, T. Zhao, P. J. Starks, D. D. Bosch, M. Seyfried, M. S. Moran, D. C. Goodrich, Y. H. Kerr, and D. Leroux, "Validation of Soil Moisture and Ocean Salinity (SMOS) Soil Moisture Over Watershed Networks in the U.S.," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1530–1543, May 2012.

[47] I. Gherboudj, R. Magagi, K. Goita, A. A. Berg, B. Toth, and A. Walker, "Validation of SMOS Data Over Agricultural and Boreal Forest Areas in Canada," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1623–1635, May 2012.

[48] R. Oliva, E. Daganzo, Y. H. Kerr, S. Mecklenburg, S. Nieto, P. Richaume, and C. Gruhier, "SMOS Radio Frequency Interference Scenario: Status and Actions Taken to Improve the RFI Environment in the 1400–1427-MHz Passive Band," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1427–1439, May 2012.

[49] S. V. Kumar, C. D. Peters-Lidard, Y. Tian, P. R. Houser, J. Geiger, S. Olden, L. Lighty, J. L. Eastman, B. Doty, P. Dirmeyer, J. Adams, K. Mitchell, E. F. Wood, and J. Sheffield, "Land information system: An interoperable framework for high resolution land surface modeling," *Environ. Model. Softw.*, vol. 21, no. 10, pp. 1402–1415, Oct. 2006.

[50] C. D. Peters-Lidard, P. R. Houser, Y. Tian, S. V. Kumar, J. Geiger, S. Olden, L. Lighty, B. Doty, P. Dirmeyer, J. Adams, K. Mitchell, E. F. Wood, and J. Sheffield, "High-performance Earth system modeling with NASA/GSFC's Land Information System," *Innov. Syst. Softw. Eng.*, vol. 3, no. 3, pp. 157–165, 2007.

[51] M. B. Ek, “Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model,” *J. Geophys. Res.*, vol. 108, no. D22, p. 8851, Nov. 2003.

[52] W. C. Skamarock and J. B. Klemp, “A time-split nonhydrostatic atmospheric model for weather research and forecasting applications,” *J. Comput. Phys.*, vol. 227, no. 7, pp. 3465–3485, Mar. 2008.

[53] Y. Xia, M. B. Ek, Y. Wu, T. W. Ford, and S. M. Quiring, “Comparison of NLDAS-2 Simulated and NASMD Observed Daily Soil Moisture. Part II: Impact of Soil Texture Classification and Vegetation Type Mismatches,” *J. Hydrometeorol.*, Jun. 2015.

[54] G. Evensen, “The Ensemble Kalman Filter: Theoretical formulation and practical implementation,” *Ocean Dyn.*, vol. 53, no. 4, pp. 343–367, 2003.

[55] R. H. Reichle, J. P. Walker, R. D. Koster, and P. R. Houser, “Extended versus Ensemble Kalman Filtering for Land Data Assimilation,” *J. Hydrometeorol.*, vol. 3, no. 6, pp. 728–740, Dec. 2002.

[56] S. V. Kumar, R. H. Reichle, C. D. Peters-Lidard, R. D. Koster, X. Zhan, W. T. Crow, J. B. Eylander, and P. R. Houser, “A land surface data assimilation framework using the land information system: Description and applications,” *Adv. Water Resour.*, vol. 31, no. 11, pp. 1419–1432, Nov. 2008.

[57] J. M. Sabater, L. Jarlan, J.-C. Calvet, F. Bouyssel, and P. De Rosnay, “From Near-Surface to Root-Zone Soil Moisture Using Different Assimilation Techniques,” *Journal of Hydrometeorology*, vol. 8, no. 2. pp. 194–206, 2007.

[58] Food and Agriculture Organization of the United Nations (FAO), “AQUASTAT website,” 2015. [Online]. Available: www.fao.org/nr/water/aquastat/irrigationmap/index.stm. [Accessed: 07-Apr-2015].

[59] S. Siebert, V. Henrich, K. Frenken, and J. Burke, “Update of the Global Map of Irrigation Areas to version 5,” *Univ. Bonn/FAO, Bonn, Ger. Italy, 178pp., doi*, vol. 10, no. 2.1, pp. 2660–6728, 2013.

[60] D. P. Dee and A. M. Da Silva, “Data assimilation in the presence of forecast bias,” *Q. J. R. Meteorol. Soc.*, vol. 124, no. 545, pp. 269–295, Jan. 1998.

[61] H. Wilker, M. Drusch, G. Seuffert, and C. Simmer, “Effects of the Near-Surface Soil Moisture Profile on the Assimilation of L-band Microwave Brightness Temperature,” *J. Hydrometeorol.*, vol. 7, no. 3, pp. 433–442, Jun. 2006.

[62] G. J. M. De Lannoy, R. H. Reichle, P. R. Houser, V. R. N. Pauwels, and N. E. C. Verhoest, "Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter," *Water Resour. Res.*, vol. 43, no. 9, Sep. 2007.

[63] D. P. Dee and S. Uppala, "Variational bias correction of satellite radiance data in the ERA-Interim reanalysis," *Q. J. R. Meteorol. Soc.*, vol. 135, no. 644, pp. 1830–1841, Oct. 2009.

[64] T. J. Jackson, R. Bindlish, M. H. Cosh, T. Zhao, P. J. Starks, D. D. Bosch, M. Seyfried, M. S. Moran, D. C. Goodrich, Y. H. Kerr, and D. Leroux, "Validation of soil moisture and Ocean Salinity (SMOS) soil moisture over watershed networks in the U.S.," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5 PART 1, pp. 1530–1543, 2012.

[65] R. D. Koster and P. C. D. Milly, "The Interplay between Transpiration and Runoff Formulations in Land Surface Schemes Used with Atmospheric Models," *J. Clim.*, vol. 10, no. 7, pp. 1578–1591, Jul. 1997.

[66] C. B. Blankenship and W. L. Crosson, "Bias correction based on regime-dependent Cumulative Distribution Functions for soil moisture data assimilation in a Land Surface Model (91st American Meteorological Society Annual Meeting)," in *AMS 25th Conf. on Hydrology*, 2011, p. 48.

[67] D. A. Miller and R. A. White, "A Conterminous United States Multilayer Soil Characteristics Dataset for Regional Climate and Hydrology Modeling," *Earth Interact.*, vol. 2, no. 2, pp. 1–26, Jan. 1998.

[68] T. R. Loveland, B. C. Reed, J. F. Brown, D. O. Ohlen, Z. Zhu, L. Yang, and J. W. Merchant, "Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data," *Int. J. Remote Sens.*, vol. 21, no. 6–7, pp. 1303–1330, Nov. 2010.

[69] S. A. Indra Espacio, "SMOS Level 2 and Auxiliary Data Products Specifications", SO-TN-IDRGS-0006, Issue 6.1, 2012.

[70] A. Al-Yaari, J.-P. Wigneron, A. Ducharne, Y. Kerr, P. de Rosnay, R. de Jeu, A. Govind, A. Al Bitar, C. Albergel, J. Muñoz-Sabater, P. Richaume, and A. Mialon, "Global-scale evaluation of two satellite-based passive microwave soil moisture datasets (SMOS and AMSR-E) with respect to Land Data Assimilation System estimates," *Remote Sens. Environ.*, vol. 149, pp. 181–195, Jun. 2014.

[71] B. T. Zavodsky, J. L. Case, C. B. Blankenship, W. L. Crosson, and K. D. White, "Application of Next-Generation Satellite Data to a High-Resolution, Real-Time Land Surface Model," *Earthzine*, 2013.

[72] K. D. White and J. L. Case, “Operational Assessment of 3-km Land Information System Soil Moisture Data for Drought Monitoring and Hydrologic Applications,” in *95th American Meteorological Society Annual Meeting*, 2015, p. 520.

[73] J. S. Kain, S. J. Weiss, D. R. Bright, M. E. Baldwin, J. J. Levit, G. W. Carbin, C. S. Schwartz, M. L. Weisman, K. K. Droegemeier, D. B. Weber, and K. W. Thomas, “Some Practical Considerations Regarding Horizontal Resolution in the First Generation of Operational Convection-Allowing NWP,” *Wea. Forecasting*, vol. 23, no. 5, pp. 931–952, Oct. 2008.

[74] S. V. Kumar, R. H. Reichle, K. W. Harrison, C. D. Peters-Lidard, S. Yatheendradas, and J. A. Santanello, “A comparison of methods for a priori bias correction in soil moisture data assimilation,” *Water Resour. Res.*, vol. 48, no. 3, Mar. 2012.

[75] D. J. Leroux, Y. H. Kerr, P. Richaume, and R. Fieuzal, “Spatial distribution and possible sources of SMOS errors at the global scale,” *Remote Sens. Environ.*, vol. 133, pp. 240–250, 2013.

[76] T. W. Collow, A. Robock, J. B. Basara, and B. G. Illston, “Evaluation of SMOS retrievals of soil moisture over the central United States with currently available in situ observations,” *J. Geophys. Res.*, vol. 117, no. D9, p. D09113, May 2012.

[77] Y. Xia, K. Mitchell, M. Ek, B. Cosgrove, J. Sheffield, L. Luo, C. Alonge, H. Wei, J. Meng, B. Livneh, Q. Duan, and D. Lohmann, “Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow,” *Journal of Geophysical Research*, vol. 117, no. D3. 2012.

[78] Y. Lin and K. E. Mitchell, “The NCEP Stage II / IV Hourly Precipitation Analyses : Development and Applications,” in *19th Conf. on Hydrology, American Meteorological Society*, 2005, pp. 2–5.

[79] M. A. Friedl, D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang, “MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets,” *Remote Sens. Environ.*, vol. 114, no. 1, pp. 168–182, Jan. 2010.

[80] B. J. Cosby, G. M. Hornberger, R. B. Clapp, and T. Ginn, “A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils,” *Water Resour. Res*, vol. 20, no. 6, pp. 682–690, 1984.

[81] J. L. Case, F. J. Lafontaine, J. R. Bell, G. J. Jedlovec, S. V. Kumar, and C. D. Peters-Lidard, “A real-time MODIS vegetation product for land surface and numerical weather prediction models,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1772–1786, Mar. 2014.

[82] S. Quiring, T. Ford, J. Wang, A. Khong, E. Harris, T. Lindgren, D. Goldberg, and Z. Li, 2015: The North American Soil Moisture Database: Development and Applications. *Bull. Amer. Meteor. Soc.* doi:10.1175/BAMS-D-13-00263.1, in press.

[83] C. Albergel, E. Zakharova, J. C. Calvet, M. Zribi, M. Pardé, J. P. Wigneron, N. Novello, Y. Kerr, A. Mialon, and N. E. D. Fritz, "A first assessment of the SMOS data in southwestern France using in situ and airborne soil moisture estimates: The CAROLS airborne campaign," *Remote Sens. Environ.*, vol. 115, no. 10, pp. 2718–2728, Oct. 2011.

[84] A. Al Bitar, D. Leroux, Y. H. Kerr, O. Merlin, P. Richaume, A. Sahoo, and E. F. Wood, "Evaluation of SMOS Soil Moisture Products Over Continental U.S. Using the SCAN/SNOTEL Network," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1572–1586, May 2012.

[85] S. Bircher, J. E. Balling, N. Skou, and Y. H. Kerr, "Validation of SMOS Brightness Temperatures During the HOBE Airborne Campaign, Western Denmark," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1468–1482, May 2012.

[86] H. Su, Z.-L. Yang, G.-Y. Niu, and C. R. Wilson, "Parameter estimation in ensemble based snow data assimilation: A synthetic study," *Adv. Water Resour.*, vol. 34, no. 3, pp. 407–416, Mar. 2011.

[87] K. Yang, T. Koike, I. Kaihotsu, and J. Qin, "Validation of a Dual-Pass Microwave Land Data Assimilation System for Estimating Surface Soil Moisture in Semiarid Regions," *J. Hydrometeorol.*, vol. 10, no. 3, pp. 780–793, Jun. 2009.

[88] W. A. Lahoz and G. J. M. De Lannoy, 2014: Closing the Gaps in Our Knowledge of the Hydrological Cycle over Land: Conceptual Problems. *Surv. Geophys.*, **35**, 623–660, doi:10.1007/s10712-013-9221-7.

[89] S. V. Kumar, C. D. Peters-Lidard, J. Santanello, K. Harrison, Y. Liu, and M. Shaw, "Land surface Verification Toolkit (LVT) – a generalized framework for land surface model evaluation," *Geosci. Model Dev.*, vol. 5, no. 3, pp. 869–886, Jun. 2012.

Aggregated Category	STATSGO Index	STATSGO Description	Percent of Domain
Sand	1	Sand	7.6
Sandy Loam Group	3	Sandy Loam	20.1
	2	Loamy Sand	4.5
Silt Loam Group	4	Silt Loam	37.5
	13	Organic	1.4
Loam	6	Loam	12.6
Clay Loam Group	8	Silty Clay Loam	5.8
	9	Clay Loam	3.6
	7	Sandy Clay Loam	0.7
Silty Clay	11	Silty Clay	1.6
Clay	12	Clay	4.6

Table 1. *Soil type category groupings used for establishing bias corrections, with corresponding STATSGO indices and percentage of the domain occupied by each type.*

Aggregated Category	IGBP Index	IGBP Description	Percent of Domain
Deciduous Forest	4	Deciduous Broadleaf Forest	18.4
Other Forest	5	Mixed Forests	10.7
	8	Woody Savannas	3.0
	2	Evergreen Broadleaf Forest	2.4
	1	Evergreen Needleleaf Forest	1.9
	3	Deciduous Needleleaf Forest	<0.1
Crops	12	Croplands	28.8
	14	Cropland/natural vegetation mosaic	16.6
	16	Barren or sparsely vegetated	<0.1
Grasslands	10	Grasslands	14.3
	9	Savannas	0.3
Shrublands	7	Open Shrublands	0.8
	6	Closed Shrublands	<0.1
Urban	13	Urban and Built-Up	2.7

Table 2. *Land use category groupings used for establishing bias corrections, with corresponding IGBP indices and percentage of the domain occupied by each type.*

Experiment Code	Name	Data Assimilation?	Bias Correction
OPL	Open Loop	No	None
NoBC	No Bias Correction	Yes	None
BC1	Uniform Bias Correction	Yes	Uniform
BCS	Soil Bias Correction	Yes	Soil-based
BCV	Landcover Bias Correction	Yes	Landcover-based

Table 3. *Overview of simulations conducted.*

Metric	Value
Bias	-0.060 ± 0.013
RMSE	0.120 ± 0.007
ubRMSE	0.071 ± 0.003
Correlation	0.553 ± 0.027
Anomaly Correlation	0.411 ± .0.24

Table 4. *Statistics for SMOS retrievals compared against ground stations.*

Metric	Upper Layer Soil Moisture				
# Stations	194				
Experiment	OPL	NoBC	BC1	BCS	BCV
Bias	**-0.000** ± 0.011	-0.026 ± 0.011	-0.023 ± 0.011	-0.005 ± 0.011	-0.025 ± 0.011
RMSE	**0.082** ± 0.005	0.087 ± 0.006	0.086 ± 0.005	**0.082** ± 0.005	0.087 ± 0.006
Unbiased RMSE	0.046 ± 0.003	**0.043** ± 0.002	**0.043** ± 0.002	0.044 ± 0.003	**0.043** ± 0.002
Correlation	0.451 ± 0.023	**0.573** ± 0.027	0.569 ± 0.026	0.539 ± 0.025	0.561 ± 0.026

Table 5. *Experimental error statistics with 95% confidence intervals for 0-10 cm layer soil moisture, verified against Texas A&M North American Soil Moisture Database in situ observations from 1 April to 1 October 2011. Experiment configurations are summarized in* Table 3. *The best statistics in each category are in bold font.*

Metric	Root Zone Soil Moisture				
# Stations	137				
Experiment	OPL	NoBC	BC1	BCS	BCV
Bias	0.038 ± 0.015	-0.013 ± 0.016	**-0.002** ± 0.016	0.014 ± 0.016	-0.009 ± 0.017
RMSE	0.093 ± 0.008	0.094 ± 0.008	**0.092** ± 0.008	**0.092** ± 0.008	0.094 ± 0.008
Unbiased RMSE	0.037 ± 0.003	0.040 ± 0.003	**0.036** ± 0.002	0.038 ± 0.003	0.038 ± 0.003
Correlation	0.672 ± 0.040	**0.685** ± 0.043	0.680 ± 0.043	0.667 ± 0.042	0.677 ± 0.045

Table 6. *As Table 5, but for root zone (top 1 m) soil moisture.*

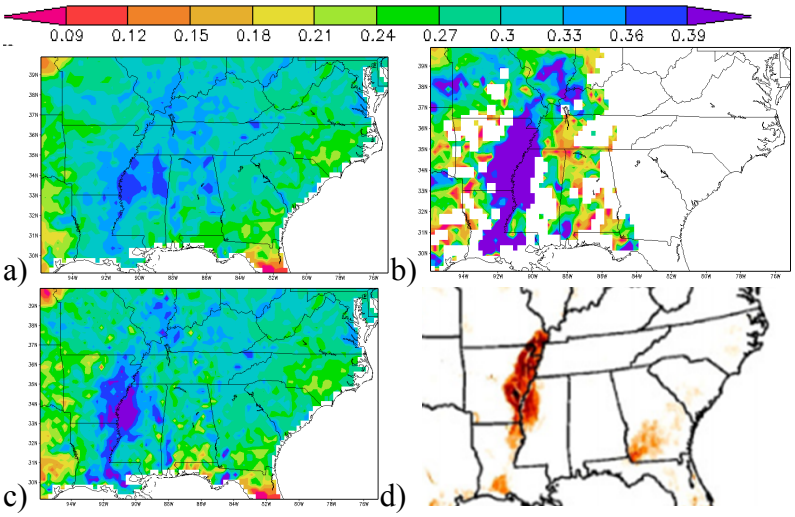


Figure 1. Soil moisture data assimilation case from 1 April 2013: a) Model background 0-10 cm volumetric soil moisture (cm3/cm3); b) SMOS soil moisture retrieval; c) 0-10 cm soil moisture analysis; d) UN Food and Agricultural Organization map of irrigated areas.

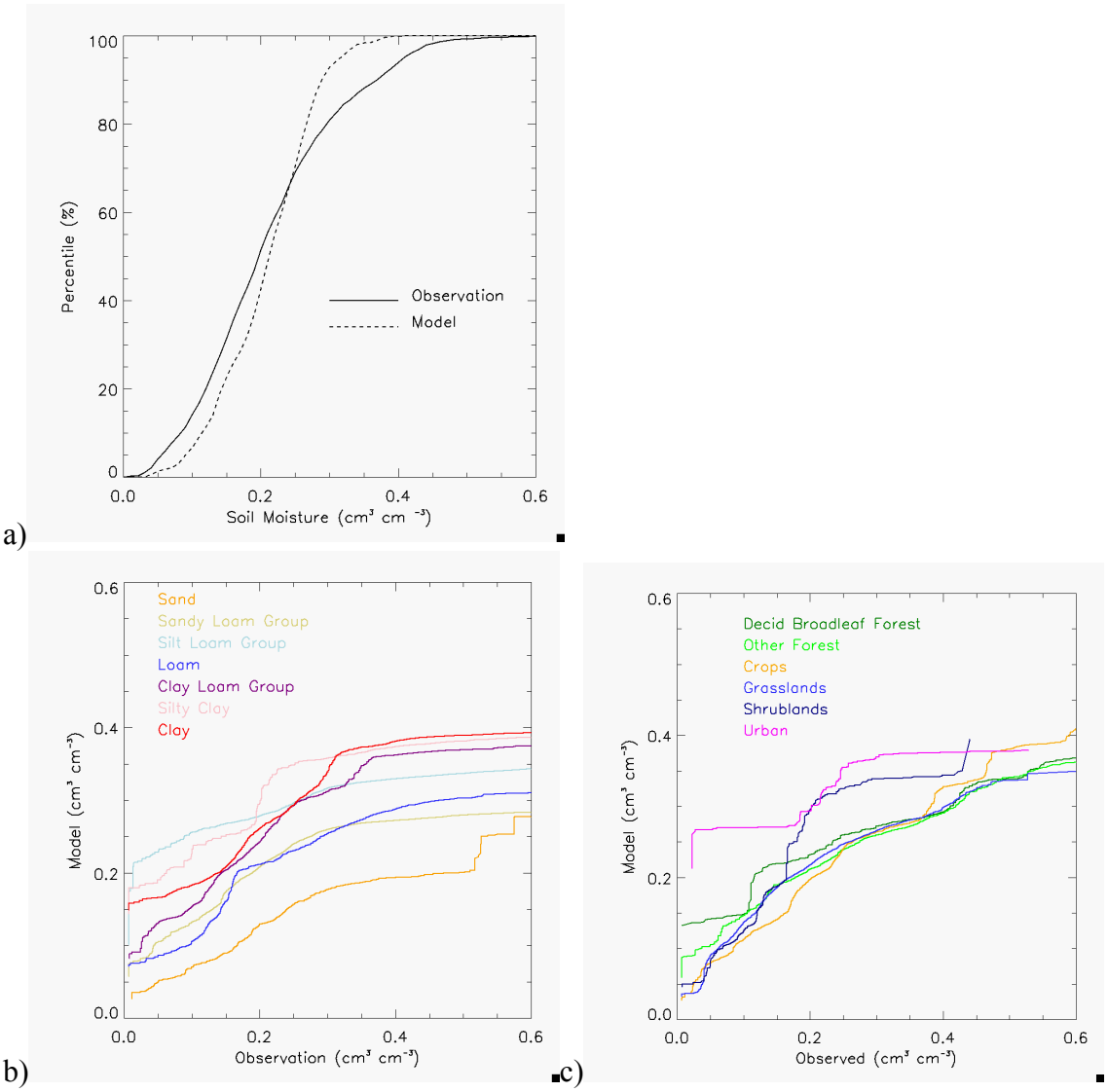


Figure 2. a) Cumulative Distribution Functions (CDFs) of SMOS retrieved surface soil moisture (solid curves) and modeled 0-10 cm soil moisture (dashed curves) for all model points, for March-September 2013. b) Correction curves derived from CDFs for six vegetation categories. c) Correction curves derived from CDFs for seven soil type categories.

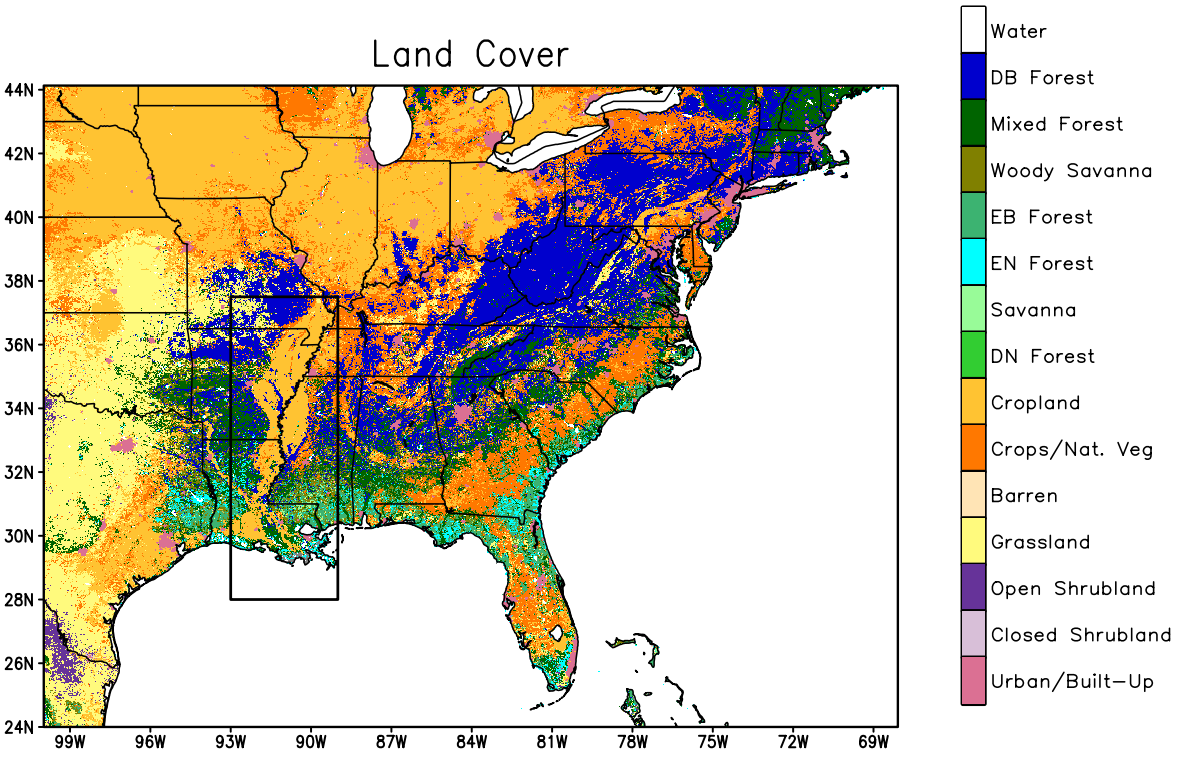


Figure 3. Land cover categories for the model domain. Rectangular box indicates area excluded from CDF matching database prior to June 1.

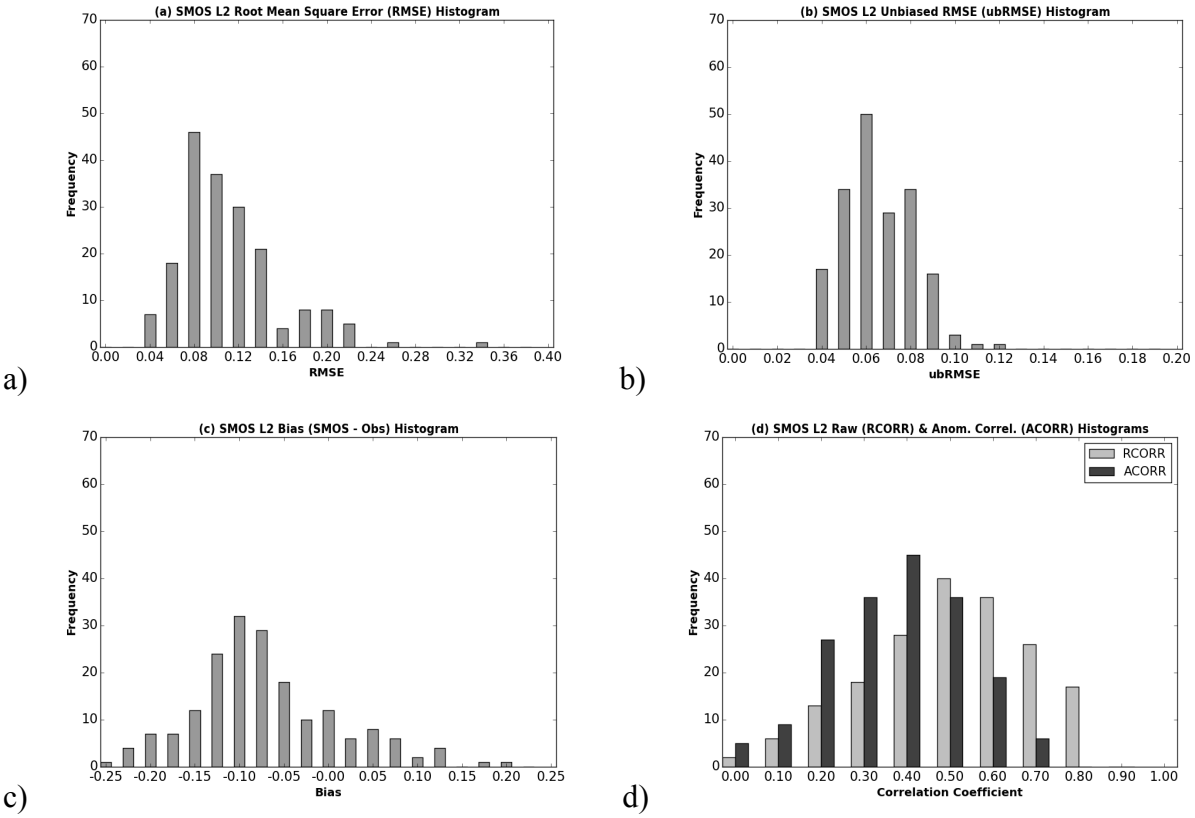
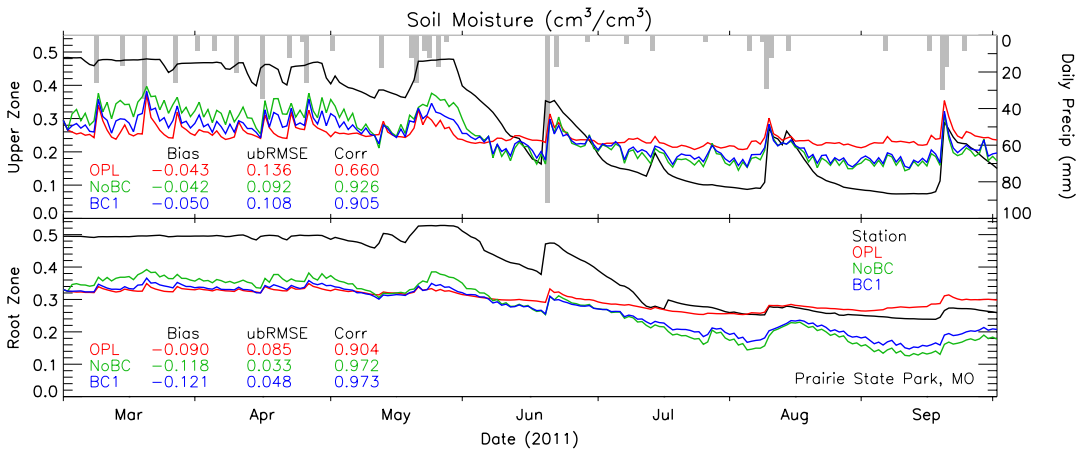
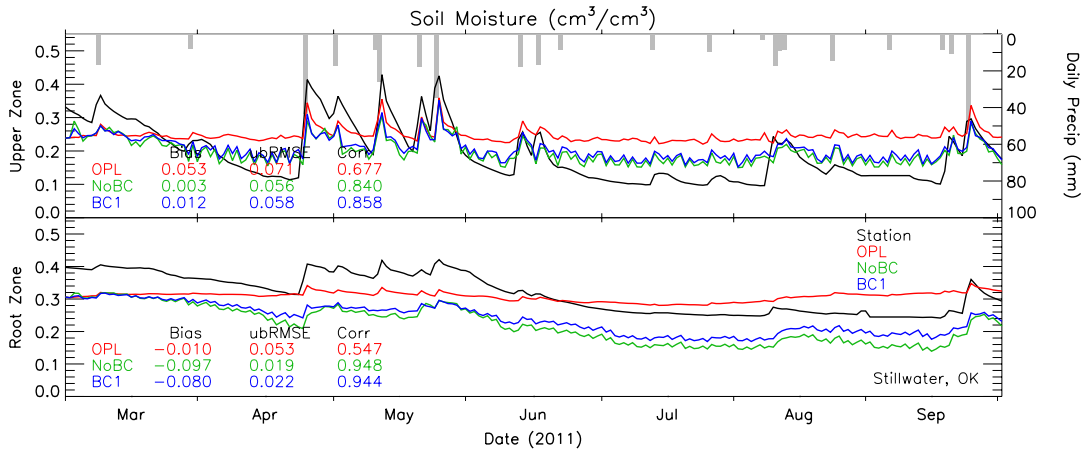


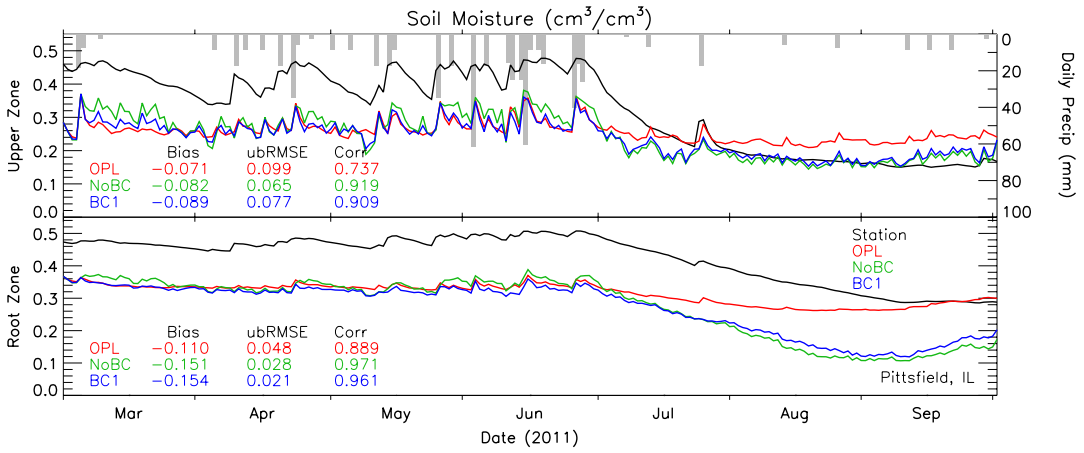
Figure 4. Histograms of (a) RMSE, (b) ubRMSE, (c) bias, and (d) correlation (RCORR) and anomaly correlation (ACORR),, for SMOS Level 2 retrieved soil moisture, as compared to NASMD ground station measurements. The distribution of values is created by calculating these metrics separately at each ground station.



a)



b)



c)

Figure 5. Time series of upper zone (0-10 cm) and root zone (0-1 m) *in situ* soil moisture (black) and collocated model soil moisture [red: open loop (OPL); green: SMOS data assimilation without bias correction (NoBC); blue: SMOS data assimilation with uniform bias correction (BC1)] from 1 Mar 2011 to 30 Sep 2011, for three locations: a) Prairie State Park, Missouri; b) Stillwater, Oklahoma; c) Pittsfield, Illinois. Vertical bars indicate daily precipitation (scale on right).

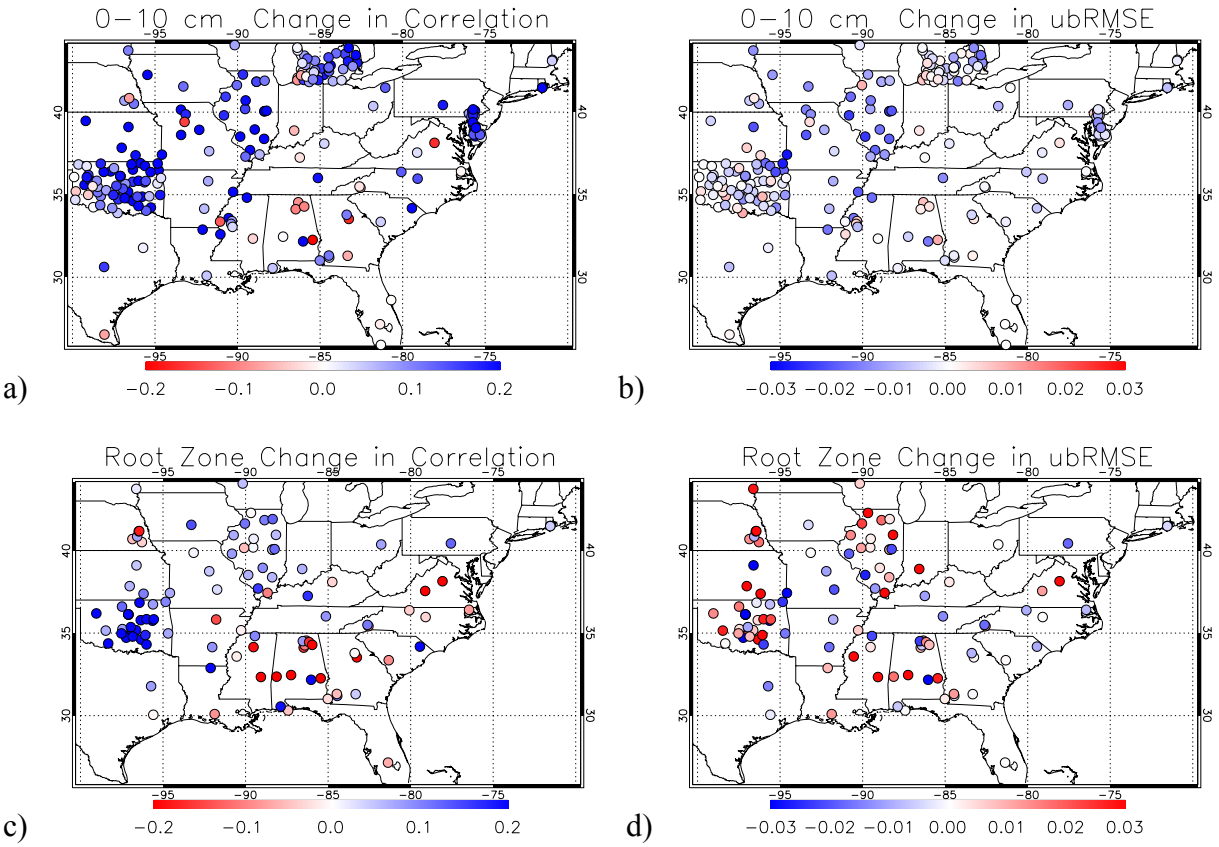


Figure 6. Changes in validation statistics (SMOS data assimilation uniform bias correction run BC1 minus open loop control run OPL) at field stations for 1 Apr 2011 to 1 October 2011. a) 0-10 cm correlation; b) 0-10 cm unbiased RMSE (cm^3/cm^3); c) root zone correlation; d) root zone unbiased RMSE (cm^3/cm^3). In each case, blue indicates improvement and red indicates degradation when compared to the station measurements.